

# DNA 末端硫修饰依赖的体外大片段 DNA 连接法

胡媚月, 吴更\*

上海交通大学生命科学技术学院 微生物代谢国家重点实验室, 上海 200030

胡媚月, 吴更. DNA 末端硫修饰依赖的体外大片段 DNA 连接法[J]. 微生物学通报, 2023, 50(3): 1220-1230.

HU Meiyue, WU Geng. Assembling long DNA fragments with DNA sulfur modification *in vitro*[J]. Microbiology China, 2023, 50(3): 1220-1230.

**摘要:**【背景】DNA 组装技术是基因组合成中的一个关键技术。探索低成本、高效率的基因组合成技术一直是合成生物学的重要研究领域。在某些细菌如变铅青链霉菌中, DNA 上有磷硫酰化修饰(简称硫修饰), 而在另一些细菌如天蓝色链霉菌中存在一种含有硫修饰识别结构域(sulfur-binding domain, SBD)的识别蛋白, 可以特异性识别 DNA 上的硫修饰, 这启发了我们发展出一种新的 DNA 组装技术。【目的】探究在 DNA 末端硫修饰的连接中, T4 DNA 连接酶与 SBD 融合蛋白和单独的 T4 DNA 连接酶相比, 是否有更高的连接效率。【方法】根据同源重组原理, 设计硫修饰引物, 扩增硫修饰的 DNA 片段。构建 T4 DNA 连接酶与 SBD 融合蛋白的 3 种表达载体 T4-linker-SBD (Hga)、T4-linker-SBD (Spr)和 T4-linker-SBD (Mmo), 表达纯化以上 3 种融合蛋白。比较 3 组浓度梯度(2.4、0.24、0.024 mg/mL) T4 DNA 连接酶与融合蛋白在 2.5 kb 和 8.0 kb DNA 片段连接上的差异。【结果】DNA 末端硫修饰的 2.5 kb 和 8.0 kb 的两端片段均能扩增, 而且 3 种融合蛋白均能表达。在两段 2.5 kb 片段的体外连接实验中, 每组 T4 DNA 连接酶与融合蛋白均能连接成 5 kb 片段; 然而在两段 8.0 kb 片段的体外连接实验中, 相同连接时间和相同酶浓度下, 仅融合蛋白 T4-linker-SBD (Hga)能够将 8.0 kb 片段连接成 16 kb 的长片段。【结论】根据两组琼脂糖凝胶电泳的实验结果可知, 在两段 8.0 kb 的 DNA 长片段连接实验中, T4-linker-SBD (Hga)融合蛋白具有更高的连接效率。本研究方法以较低的成本和比较高的效率为大片段 DNA 组装提供了一种新的方法和思路。

**关键词:** 硫原子识别结构域; DNA 连接组装; DNA 硫修饰; 基因组合成

资助项目: 国家自然科学基金(31670106)

This work was supported by the National Natural Science Foundation of China (31670106).

\*Corresponding author. E-mail: geng.wu@sjtu.edu.cn

Received: 2022-06-19; Accepted: 2022-07-29; Published online: 2022-11-11

# Assembling long DNA fragments with DNA sulfur modification *in vitro*

HU Meiyue, WU Geng\*

State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200030, China

**Abstract:** **[Background]** Assembling large DNA fragments is crucial in genome synthesis. Exploring inexpensive and efficient DNA assembly technologies has been a research hotspot in synthetic biology. In some bacteria such as *Streptomyces lividans*, there exists phosphorothioation modification on DNA (also known as DNA sulfur modification), while in some other bacteria such as *Streptomyces coelicolor*, there is a protein possessing sulfur-binding domain (SBD), which serves as a reader protein to specifically recognize the sulfur modification on DNA. This inspired us to develop a new technology to ligate long DNA fragments. **[Objective]** In this study, we introduced sulfur modification at the 5' and 3' ends of the DNA fragments to be ligated, and employed a fusion protein containing SBD domain and T4 DNA ligase to complete the ligation of long DNA fragments *in vitro* and compare the different efficiency between 2.5 kb and 8 kb fragments with these two ligases. **[Methods]** We designed sulfur-modified primers and amplified 2.5 kb and 8 kb DNA fragments with sulfur modification. Additionally, we constructed three fusion protein vectors: T4-linker-SBD (Hga), T4-linker-SBD (Spr), and T4-linker-SBD (Mmo) and expressed them in *Escherichia coli* and compared the ligation efficiency between T4 DNA ligase and fusion proteins under three different concentration gradients: 2.4, 0.24, 0.024 mg/mL. **[Results]** We successfully amplified 2.5 kb and 8 kb DNA fragments with sulfur modification and expressed these three fusion proteins. In the 2.5 kb DNA fragments ligation experiment, we found that T4 DNA ligase and fusion protein under the three concentration gradients and in twelve groups all successfully ligated, whereas in the 8 kb DNA fragments ligation experiment, only T4-linker-SBD (Hga) fusion protein successfully ligated two 8 kb fragments into 16 kb fragment. **[Conclusion]** According to our agarose gel electrophoresis, we concluded that in the 8 kb DNA fragments ligation experiment, T4-linker-SBD (Hga) fusion protein had higher ligation efficiency. Thus our method provide a new way to assemble large DNA fragments with low cost and high efficiency.

**Keywords:** sulfur-binding domain (SBD); DNA ligation and assembly; DNA sulfur modification; genome synthesis

合成人造细胞一直是合成生物学领域一个重要的研究内容。合成人造细胞的一个关键环节就是基因组的合成，而 DNA 组装技术是基因组合成的一个关键技术。随着 DNA 测序技术的发展，越来越多物种的基因组序列已经被

测序和解读。在过去几十年里，将短片段 DNA 组装成长片段 DNA 的组装技术和方法在不断地发展。最具有代表性的序列依赖型重组技术就是 Gibson<sup>[1-3]</sup> 组装，这也是目前主流的长片段 DNA 组装技术。利用 T5 核酸外切酶在

序列末端形成互补配对的黏性末端，继而根据同源重组的原理，将 DNA 片段组装起来；最后利用 DNA 聚合酶补上缺口，*Taq* DNA 连接酶补上切刻，形成完整的 DNA 片段。Gibson 等<sup>[1]</sup>利用 Gibson 组装法体外组装了生殖支原体 1/4 的基因组(144 kb)，他们将支原体基因组分成若干个 2 kb 的 DNA 片段，组装成 25 个 24 kb 的 DNA 大片段，再将 3 个 24 kb 的 DNA 片段组装成一个 72 kb 的 DNA 大片段，每 2 个 72 kb 的 DNA 片段组装成一个 144 kb 的 DNA 大片段；随着载体插入片段越来越大，转入大肠杆菌的效率越来越低，之后体外组装难以进行下去，最后将这些 144 kb 的 DNA 大片段在酵母细胞的重组系统中组装成一个完整的合成基因组(583 kb)。随后，Gibson 等<sup>[3]</sup>利用酵母组装系统完成了 1.08 Mb 蕈状支原体基因组的组装，并移植到另一个去基因组的支原体中，发现细胞可以生长繁殖。

然而，Gibson 方法需要在酵母中通过同源重组进行连接，是一种依赖于天然生命体的 DNA 体内连接方法，因此不是完全意义上的“从头合成人造生命”。而且，随着生物体复杂性的提高，基因组越来越大。例如，酵母、果蝇、小鼠的基因组分别为 12、147、2 700 Mb ([https://www.ncbi.nlm.nih.gov/genome/15?genome\\_assembly\\_id=22535](https://www.ncbi.nlm.nih.gov/genome/15?genome_assembly_id=22535)、[https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA\\_016746445.1/](https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_016746445.1/)、[https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF\\_015227675.2/](https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_015227675.2/))。组装或合成如此大的全基因组对于 Gibson 法等通过体内同源重组来不断拼接小片段 DNA 的组装方法带来了极大的挑战，亟须发展出一种不依赖于酵母等天然生命体、完全采用体外生化反应、易于标准化和流程化的大片段 DNA 连接方法。

DNA 作为遗传信息的载体，其化学结构

包含了 C、H、O、N、P 这 5 种基本的元素。在 2005 年，Zhou 等<sup>[4]</sup>发现细菌中天然的 DNA 硫修饰，这是一种 DNA 骨架上的表观遗传修饰，把 DNA 磷酸二酯键上的一个非桥联氧原子替换成硫原子<sup>[5]</sup>。类似于 DNA 甲基化修饰中 UHRF1 蛋白通过 SET and RING-associated (SRA)结构域识别半甲基化位点来将 DNA 甲基转移酶 1 (Dnmt1)引导到这些位点<sup>[6-13]</sup>，这类 DNA 骨架上的硫修饰也依赖于一类特定的蛋白质去识别。Liu 等<sup>[14]</sup>在天蓝色链霉菌中发现了一个识别 DNA 修饰的 Reader 蛋白 ScoMcrA，研究发现 ScoMcrA 中包含一段硫原子识别结构域(sulfur-binding domain, SBD)，该结构域将一个带正电的 loop 插进 DNA 分子的大沟里，识别 DNA 上特定的碱基序列(GpsGCC)，并将硫原子包含在该结构域的一个疏水表面空腔里。Yu 等<sup>[15]</sup>发现来源于始旋链霉菌的 SBD 序列能够识别更多天然硫核心序列(GGCC、GATC、GAAC)，其分子机制是硫原子附近的无规则卷曲结构能够与 DNA 磷酸骨架产生静电作用；当把天蓝色链霉菌 SBD 序列的无规则卷曲的关键位点所带的负电荷转变成正电荷时，也可以识别更多的 DNA 硫修饰序列。

基于以上研究内容设计硫修饰的引物，经过 PCR 扩增出末端带有硫修饰的 DNA 片段，用 T7 核酸外切酶进行酶切，而 T7 核酸外切酶的外切酶活性会被磷硫酰键抑制，能够精准地控制黏性末端的序列及切割的长度，形成 DNA 带有硫修饰的末端，不需要加入 DNA 聚合酶，采用 T4 DNA 连接酶即可将两段 DNA 的 5'硫修饰的磷酸末端与 3'羟基末端形成磷酸二酯键，利用 SBD 蛋白招募 DNA 连接酶，可进一步提高 DNA 片段体外组装的连接效率(图 1)，从而实现 DNA 片段的无缝连接。此外，

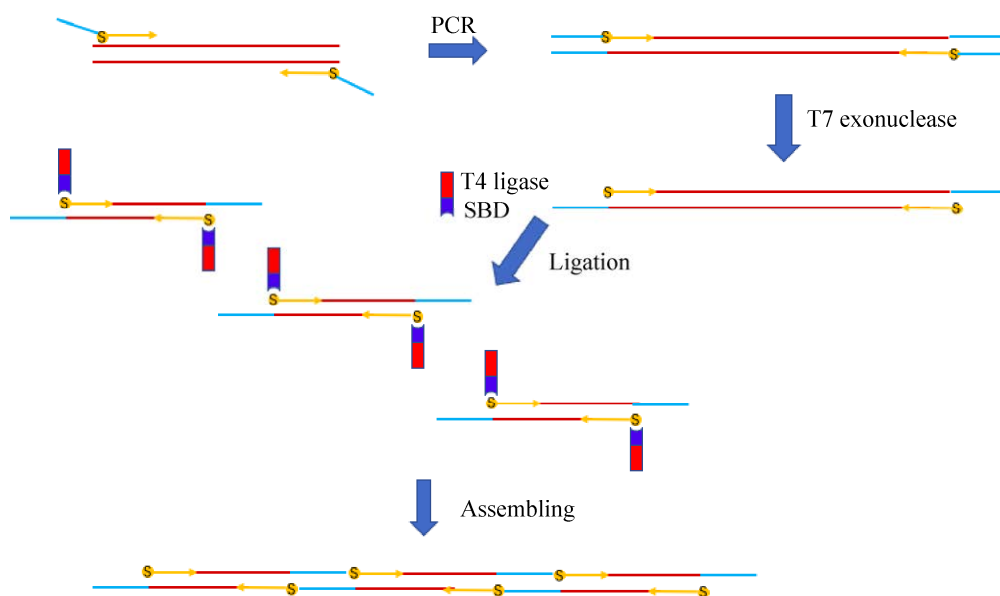


图 1 DNA 末端硫修饰依赖的体外大片段 DNA 连接法

Figure 1 The method of assembling long DNA fragments with DNA sulfur-modification dependent ligation *in vitro*.

对硫修饰有不同亲和力和序列选择性的SBD蛋白通过不同的融合方式与连接酶融合，或许可以定向提高连接的效率。目前，基因组的合成离不开生物体的体内组装，尚且算不上完全意义的合成，而体外长片段的组装尚无有效的方法，本研究提供一种长片段 DNA 体外组装的新思路和有效的连接方式，以期为体外合成基因组提供可能。

## 1 材料与方 法

### 1.1 材料

#### 1.1.1 菌株和质粒

菌株 *Escherichia coli* DH5 $\alpha$  和 *E. coli* BL21(DE3)均为本实验室保存。引物 Depdc5-1-F-2.5kb、Depdc5-1-R-2.5kb、Depdc5-2-F-2.5kb、Depdc5-2-R-2.5kb、Depdc5-1-F-8kb、Depdc5-1-R-8kb、Depdc5-2-F-8kb 和 Depdc5-2-R-8kb，华大基因公司合成。质粒 Amp-pRK5-Depdc5、Kana-pET-28a 为本实验室保存。

#### 1.1.2 主要试剂和仪器

*Taq* 酶，近岸蛋白科技有限公司；*FastPfu* 酶，全式金生物公司；核酸外切酶 T7，New England Biolabs 公司；质粒提取试剂盒和 DNA 胶回收试剂盒，捷瑞生物工程有限公司。

镍柱亲和层析平衡液：25 mmol/L pH 8.0 Tris-HCl，20 mmol/L pH 8.0 咪唑，150 mmol/L NaCl。镍柱洗脱液-梯度 1：40 mmol/L 咪唑，150 mmol/L NaCl；梯度 2：60 mmol/L 咪唑，150 mmol/L NaCl；梯度 3：110 mmol/L 咪唑，150 mmol/L NaCl；梯度 4：200 mmol/L 咪唑，150 mmol/L NaCl；梯度 5：300 mmol/L 咪唑，150 mmol/L NaCl。分子筛洗脱液：25 mmol/L pH 8.0 Tris-HCl，150 mmol/L NaCl，2 mmol/L DDT。IPTG：IPTG 固体粉末 2.38 g，超纯水 50 mL。考马斯亮蓝：5 $\times$ 考马斯亮蓝 200  $\mu$ L，超纯水 800  $\mu$ L。50 mg/mL 卡那霉素：称取 2.5 g 卡那霉素粉末，加蒸馏水至 50 mL，0.22  $\mu$ m 滤膜过滤后分装，置-20  $^{\circ}$ C 保存。1 mol/L DTT：

称取 7.71 g DTT 粉末,溶于蒸馏水中,定容至 50 mL。200 mmol/L PMSF:称量 3.48 g PMSF 固体粉末,无水乙醇溶解,定容至 250 mL。LB 培养基:25 g LB 肉汤,蒸馏水溶解,定容至 1 L, 121 °C 灭菌 30 min。

层析柱及柱材:镍离子亲和层析柱填料, Qiagen 公司;凝胶过滤层析柱 Hiload™ Superdex 200, GE Healthcare 公司;PCR 仪,朗基公司;电泳仪,韦克斯科技有限公司;凝胶成像分析系统, Tanon 公司;高压破碎仪,永联生物科技有限公司;ÄKTA™ prime Plus, GE Healthcare 公司;冷冻离心机,上海离心机械研究所;高速冷冻离心机, Beckman Coulter 公司;三孔电热恒温水槽,上海齐欣科仪公司;可见光分光光度计,上海元析仪器公司;Nano Drop 分光光度计, ThermoFisher 公司。

## 1.2 方法

### 1.2.1 T4 DNA 连接酶与 SBD 结构域蛋白融合蛋白的表达质粒构建

选取了分别来自港湾河氏菌(*Hahella ganghwensis*, Hga)、始旋链霉菌(*Streptomyces pristinaespiralis*, Spr)和摩氏摩根菌(*Morganella morganii*, Mmo)这 3 种菌种的 SBD 蛋白序列与 T4 DNA 连接酶的蛋白序列,用一段柔性 linker (氨基酸序列为 5'-GGGGSGGGGS-3')连接。采用 *Nde* I 和 *Hind* III 作为酶切位点, Kana-pET-28a 质粒作为载体进行表达载体的构建。分别按照从 N 端到 C 端构建 T4-linker-SBD (Hga)、T4-linker-SBD (Spr)、T4-linker-SBD (Mmo)这 3 个融合蛋白[以下分别简称 T4-SBD (Hga)、T4-SBD (Spr)、T4-SBD (Mmo)],在大肠杆菌原核表达系统中尝试进行表达。

### 1.2.2 融合蛋白质粒的扩增

将构建好的重组质粒进行扩增。取 1–2 μL 质粒加入到 *E. coli* DH5α 感受态细胞中,冰上孵育 20–30 min,在 42 °C 水浴中热激 90 s,冰

上孵育 3–5 min,加入 700 μL LB 培养基,在 37 °C 摇床中,以 180 r/min 的转速摇 45–60 min。将培养起来的大肠杆菌 4 000 r/min 离心 5 min,弃去 600 μL 上清,重悬菌液,涂布于卡那霉素抗性的平板,于 37 °C 培养箱中培养过夜。挑取单菌落于 50 mL LB 培养基中,加入 50 μL 卡那霉素(50 mg/mL),于 37 °C 培养箱中培养过夜,转速为 180 r/min。将培养过夜的菌液用捷瑞公司的质粒提取试剂盒提取质粒,最后加 30 μL elution buffer 洗脱目标质粒。

### 1.2.3 融合蛋白的表达与纯化

#### 1) 融合蛋白表达

将扩增的质粒取 1–2 μL,加入到 *E. coli* BL21(DE3)感受态细胞中进行转化实验,将涂布好的平板于 37 °C 培养箱中培养过夜。挑单菌落于 50 mL LB 小瓶中,于 37 °C、180 r/min 培养过夜至  $OD_{600}$  为 0.6 左右。将过夜培养的 50 mL 菌液按 1:40 倒入 2 个 1 L 的含卡那霉素(50 mg/mL)的 LB 大瓶中,于 37 °C、220 r/min 培养 4 h,将温度降低至 16 °C,180 r/min 转速下培养 2 h,加入 IPTG 诱导,培养过夜。

#### 2) 收菌

将 LB 培养基中培养过夜的菌液在 4 500×g 转速下离心 20 min,弃去上清液,加入 15–20 mL 的镍柱缓冲液,于涡旋机上重悬,重悬后转移至 50 mL 离心管中,加入 250 μL 苯甲磺酰氟 PMSF。

#### 3) 破碎收蛋白

将重悬菌液在高压破碎仪中破碎 3 遍,至流出液较为澄清。在高速冷冻离心机中,于 17 000×g 转速下离心 45 min。

#### 4) 蛋白纯化

将高速离心的上清液过两遍镍柱,用镍柱缓冲液洗脱 2–3 个柱体积,除去杂蛋白,再用不同浓度梯度的咪唑洗脱液(40、60、110、200、

300、500 mmol/L)将目的蛋白洗脱下来。取 20  $\mu$ L 蛋白洗脱液, 加入 10  $\mu$ L loading buffer, 用 15% SDS-PAGE 凝胶鉴定纯化的蛋白, 选取含目的蛋白对应的管数, 将其合并, 在 4 000 $\times$ g 转速下离心 20 min 浓缩至 2 mL。

将浓缩后的蛋白用 Superdex 200 凝胶过滤层析柱纯化, 收集出峰位置的蛋白, 用 15% SDS-PAGE 凝胶鉴定纯化的蛋白, 收集含目的蛋白的管数, 进行合并。

#### 5) 蛋白浓度测定

用 Nano Drop 测量纯化 4 种蛋白的浓度, 分别为 2.4 mg/mL T4 DNA 连接酶、5.1 mg/mL T4-SBD (Hga)融合蛋白、8.8 mg/mL T4-SBD (Mmo)融合蛋白、4.4 mg/mL T4-SBD (Spr)融合蛋白。将以上蛋白浓度稀释为 2.4 mg/mL 左右, 并以每 10  $\mu$ L 为一管分装, 冻存于-80  $^{\circ}$ C 超低温冰箱, 备用。

#### 1.2.4 硫修饰引物的设计

选取实验室保存的 Amp-pRK5-Depdc5 质粒的基因序列作为扩增模板, 长度为 9 464 bp, 利用 SnapGene 和 Oligo 软件设计硫修饰引物。利用 SnapGene 软件按照图 2 设计引物, 确定扩增片段的长度, 用 Oligo 软件来评估设计引物

是否有发夹结构和退火温度。长片段 DNA 的 PCR 扩增相对短片段更困难, 产量更低。特别是扩增 8 000 bp 的长片段 DNA 时, 为了能观察到实验结果, 提高扩增产物的浓度, 对引物进行了精细设计: (1) 引物中需要包含一段与模板互补的序列, 根据是否有发夹结构和退火温度进行长度调整; (2) 一段硫修饰序列中至少有 6 个碱基被硫修饰, 使其能够被 SBD 识别; (3) 引物中包含一段可以被外切酶剪切、形成重组互补配对黏性末端的序列。首先, 我们在 Depdc5 质粒上选取 1 个无发夹结构或解链温度较低的起始引物作为第 1 个片段 Depdc5-1 的 5'端引物, 由 10 bp 可以被外切酶切去的前端碱基(图 2 中黄色部分)+6 bp 硫修饰碱基(图 2 中红色部分)+18 bp 互补碱基(图 2 中橙色部分)组成, 总计 34 bp。第 1 个片段 Depdc5-1 的 3'端引物选取与其 5'端退火温度相近的序列, 同时没有发夹结构或解链温度低的序列, 由 10 bp 可以被外切酶切去的前端碱基+6 bp 硫修饰碱基+17 bp 互补碱基组成, 总计 33 bp。根据同源重组原理, 设计第 2 个片段 Depdc5-2 序列的 5'段引物, 有 10 bp 碱基互补, 由 10 bp 可以被外切酶切去的前端碱基+6 bp 硫修饰碱

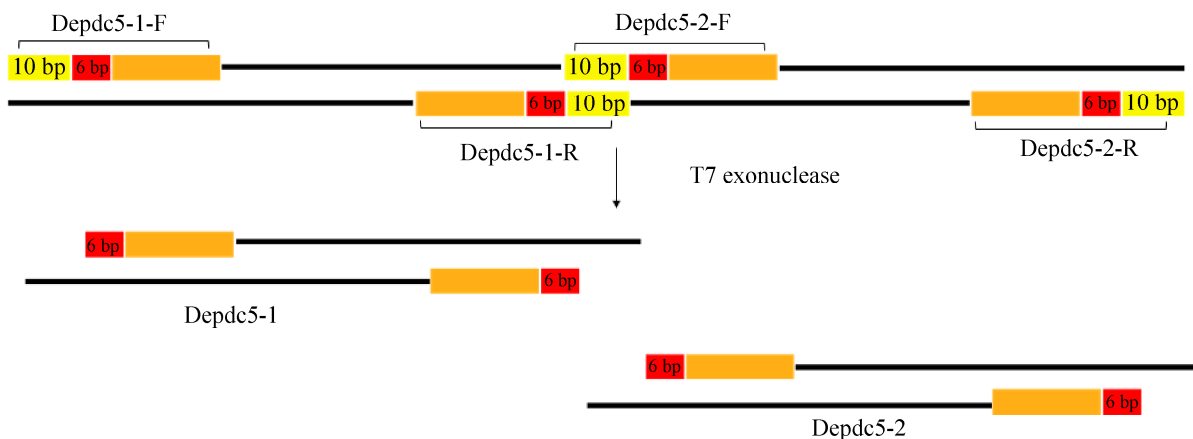


图 2 Depdc5-1-R 与 Depdc5-2-F 引物设计示意图

Figure 2 The primer design of Depdc5-1-R and Depdc5-2-F.

基+17 bp 互补碱基组成, 总计 33 bp。第 2 个片段 Depdc5-2 序列的 3'端引物设计与其 5'端退火温度相近的序列, 由 10 bp 可以被外切酶切去的前端碱基+6 bp 硫修饰碱基+19 bp 互补碱基组成, 总计 35 bp。表 1 和表 2 分别是设计的两段扩增 2 500 bp 片段的引物和两段扩增 8 000 bp 片段的引物。

### 1.2.5 长片段扩增

选取 *Pfu* 酶进行 PCR 扩增, 扩增速率为 2 kb/min。PCR 反应体系 (50  $\mu$ L): 模板 (70 ng/mL) 0.5  $\mu$ L, 5'端引物 (10  $\mu$ mol/L) 1.0  $\mu$ L, 3'端引物 (10  $\mu$ mol/L) 1.0  $\mu$ L, 硫酸镁 (50 mmol/L) 3  $\mu$ L, dNTPs (2.5 mmol/L) 4  $\mu$ L, *Pfu* 酶 (2.5  $\mu$ / $\mu$ L) 1  $\mu$ L, 5 $\times$ *Pfu* buffer 10  $\mu$ L, 超纯水 19.5  $\mu$ L, stimulator 10  $\mu$ L。PCR 反应条件: 95  $^{\circ}$ C 5 min; 95  $^{\circ}$ C 30 s, 58  $^{\circ}$ C/63  $^{\circ}$ C 30 s, 72  $^{\circ}$ C 4 min 5 s, 35 个循环; 16  $^{\circ}$ C 储存。PCR 产物用 1%琼脂糖凝胶电泳 40 min, 每两孔回收一管, 用捷瑞的 DNA 凝胶回收试剂盒进行回收, 并测量 PCR

扩增出的片段浓度为 Depdc-1-8kb 81 ng/mL (体积 60  $\mu$ L)、Depdc-2-8kb 78 ng/mL (体积 60  $\mu$ L)。

### 1.2.6 长片段 DNA 的体外连接

将 T4 DNA 连接酶、T4-SBD (Hga)、T4-SBD (Spr)、T4-SBD (Mmo)这 4 种酶依次稀释 10 倍、100 倍, 分成 3 个梯度。将扩增回收好的 Depdc5-1 片段和 Depdc5-2 片段用 T7 外切酶于 25  $^{\circ}$ C 水浴酶切 30 min。酶切体系 (25  $\mu$ L): Depdc5-1/Depdc5-2 10  $\mu$ L, T7 外切酶 0.5  $\mu$ L, T7 buffer 2.5  $\mu$ L, 超纯水 12  $\mu$ L。加入 5  $\mu$ L 90 mmol/L 的 EDTA 终止酶切后进行 1%琼脂糖凝胶电泳, 用捷瑞胶回收试剂盒回收酶切产物, 测量回收片段浓度为 Depdc-1-8kb 17 ng/mL (体积 60  $\mu$ L)、Depdc-2-8kb 14 ng/mL (体积 60  $\mu$ L)。取 4  $\mu$ L Depdc5-1-8kb DNA、4  $\mu$ L Depdc5-2-8kb DNA、1  $\mu$ L 连接酶、1  $\mu$ L T4 buffer, 在 16  $^{\circ}$ C 水浴连接过夜。连接反应共分 12 组, 分别是: T4 连接酶原浓度、T4 连接酶稀释 10 倍、T4 连接酶稀释 100 倍、T4-SBD (Hga)原浓度、T4-SBD

表 1 2 500 bp 片段的引物设计

Table 1 The primer design of 2 500 bp fragments

引物名称 Primer name	引物序列(5'→3') Sequence of primers	退火温度 Anneal temperature ( $^{\circ}$ C)
Depdc5-1-F-2.5 kb	TGCGTCGACG*ATGAGA*ACAACAAAGGTCTACAAAC	63
Depdc5-1-R-2.5 kb	AACTCTTCAA*ATACCT*GCTGGGCTGTCATCTGCAC	
Depdc5-2-F-2.5 kb	TTGAAGAGTT*TATTTG*CCAACGCTCATGCAGGGC	58
Depdc5-2-R-2.5 kb	TAGCGGCCGC*TCACGG*GGCACTGGCATGCATCTTC	

\*: 星号之间的碱基为硫修饰碱基。下同

\*: The bases between the asterisks are sulfur-modified bases. The same below.

表 2 8 000 bp 片段的引物设计

Table 2 The primer design of 8 000 bp fragments

引物名称 Primer name	引物序列(5'→3') Sequence of primers	退火温度 Anneal temperature ( $^{\circ}$ C)
Depdc5-1-F-8 kb	TCAATAATGAC*GTATGT*TCCCATAGTAACGCCAA	58
Depdc5-1-R-8 kb	TCACCGGCTC*CAGATT*TATCAGCAATAAACCCAG	
Depdc5-2-F-8 kb	GAGCCGGTGA*GCGTGG*GTCTCGCGGTATCATTG	63
Depdc5-2-R-8 kb	ATTAAAGAAC*GTGGAC*TCCAACGTCAAAGGGCGAA	

(Hga)稀释 10 倍、T4-SBD (Hga)稀释 100 倍、T4-SBD (Spr)原浓度、T4-SBD (Spr)稀释 10 倍、T4-SBD (Spr)稀释 100 倍、T4-SBD (Mmo)原浓度、T4-SBD (Mmo)稀释 10 倍、T4-SBD (Mmo)稀释 100 倍。将连接产物进行 1%琼脂糖凝胶电泳, 观察连接结果。

## 2 结果与分析

### 2.1 融合蛋白表达结果

将分子筛纯化出来的 4 个蛋白, 分别进行

15%的 SDS-PAGE 凝胶电泳。由图 3A 可见, 对照组 T4 DNA 连接酶成功表达出来, 分子量为 53.7 kDa, 用 Nano Drop 测量其蛋白浓度为 2.4 mg/mL。T4-SBD (Hga)、T4-SBD (Mmo)、T4-SBD (Spr)这 3 个融合蛋白也均在 *E. coli* BL21(DE3)中成功地诱导表达出相应的蛋白。图 3B–3D 依次为 T4-SBD (Hga)、T4-SBD (Mmo)、T4-SBD (Spr)的表达结果, 分子量分别为 72.6、73.8、73.2 kDa, 测量浓度分别为 5.1、8.8、4.4 mg/mL。

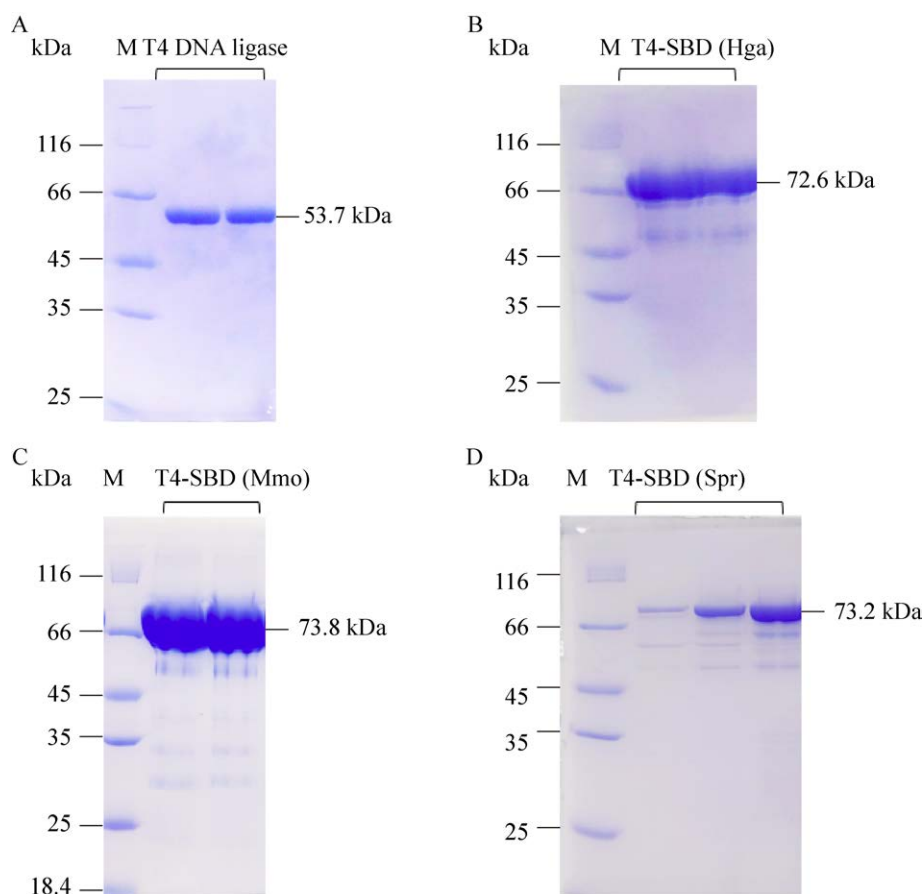


图3 各种融合蛋白的表达纯化结果 A: T4 DNA 连接酶的分子筛 Superdex 200 纯化结果. B: T4-SBD (Hga)的分子筛 Superdex 200 纯化结果. C: T4-SBD (Mmo)的分子筛 Superdex 200 纯化结果. D: T4-SBD (Spr)的分子筛 Superdex 200 纯化结果

Figure 3 The expression of various fusion proteins. A: The results of T4 ligase passing through the Superdex 200. B: The results of T4-SBD (Hga) fusion protein passing through the Superdex 200. C: The results of T4-SBD (Mmo) fusion protein passing through the Superdex 200. D: The results of T4-SBD (Spr) fusion protein passing through the Superdex 200.



## 2.2 长片段 DNA 连接

以 Depdc5 质粒为模板, 分别设计了扩增 2.5 kb 和 8 kb 的 DNA 片段的硫修饰引物, PCR 扩增了 depdc5-1-2.5 kb、depdc5-2-2.5 kb (图 4A) 和 depdc5-1-8 kb、depdc5-2-8 kb (图 4B) 这 4 个片段。

## 2.3 2.5 kb 段连接

将两段 2.5 kb 片段 DNA 以 4 组不同的酶每种酶 3 个浓度梯度进行酶连, 1% 琼脂糖凝胶电泳结果如图 5 所示, 可以观察到每组片段均可进行连接, 形成 2.5 kb 片段长度的连接产物。在相对短的片段连接中, 并未体现 T4 DNA 连

接酶与 T4-SBD 融合蛋白连接效果的差异。因此, 我们设计了扩增更长片段的片段进行连接, 观察连接效果的差异。

## 2.4 8 kb 段连接

将两段 8 kb 片段 PCR 扩增产物 DNA, 以 T4 DNA 连接酶、T4-SBD (Hga)、T4-SBD (Mmo)、T4-SBD (Spr) 这 4 组不同的酶每种酶 3 个浓度梯度进行酶连, 进行 1% 琼脂糖凝胶电泳, 以 T4 DNA 连接酶作为对照组。由连接结果(图 6)可以观察到, T4 连接酶未将 2 个 8 kb DNA 片段连接, 而 T4-SBD (Hga) 融合蛋白连接酶可以把 2 个 8 kb DNA 片段连接。我们进行了

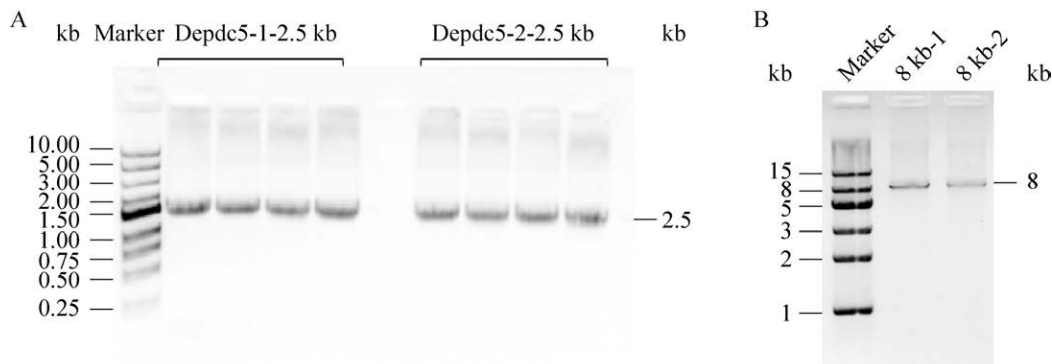


图 4 PCR 扩增产物 A: 2.5 kb 片段 PCR 扩增产物. B: 8 kb 片段 PCR 扩增产物

Figure 4 The results of PCR amplification. A: The products of 2.5 kb fragments PCR amplification. B: The products of 8 kb fragments PCR amplification.

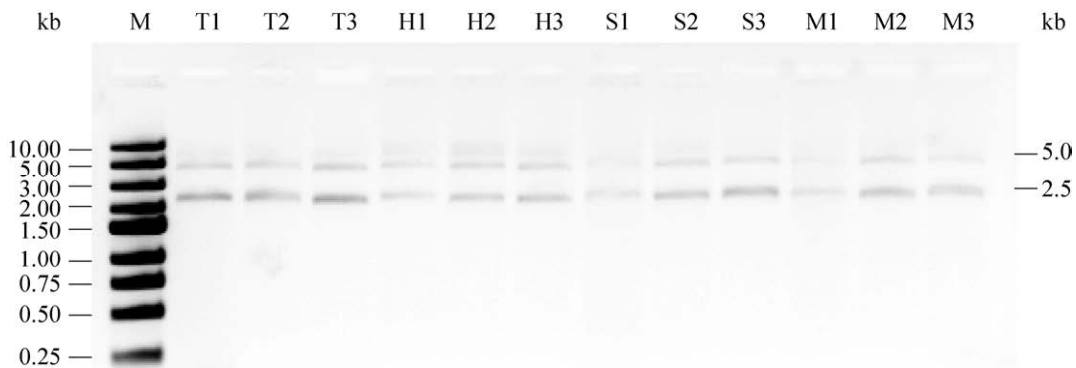


图 5 2.5 kb 片段连接成 5 kb 产物

Figure 5 The result of 2.5 kb fragments being ligated into 5 kb fragments.

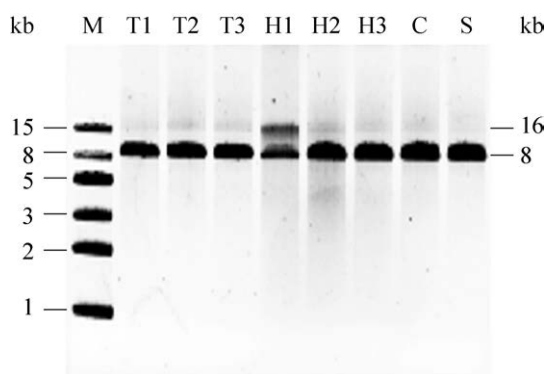


图 6 8 kb 片段 DNA 的连接结果

Figure 6 The results of 8 kb fragment DNA ligation.

3 次重复实验, T4-SBD (Hga)融合蛋白连接酶成功连接了 2 次,而 T4 连接酶 3 次均未将两个 8 kb DNA 片段连接起来。

### 3 讨论与展望

本研究成功构建了 3 种表达不同 T4-SBD 融合蛋白的质粒,并表达出了相应的融合蛋白 T4-SBD (Hga)、T4-SBD (Spr)、T4-SBD (Mmo)。我们分别设计了扩增 2.5 kb 和 8 kb 长度片段 DNA 的硫修饰引物,摸索了克隆条件,通过尽量提高回收的片段浓度,以便于后续实验步骤中能够观测到连接结果。我们通过将两组片段进行连接,比较出 T4-SBD (Hga)蛋白在长片段 DNA 组装中有较好的连接优势,而 T4 DNA 连接酶在相同的连接时间里未能将 8 kb 的 DNA 片段连接。

本文提出的硫修饰依赖的 DNA 连接方法:通过将待合成的长 DNA 片段或基因组分成若干个相等或不相等的片段,根据基因重组原理,在每个片段上设计一对硫修饰引物,经过 PCR 扩增出各个 DNA 末端带有硫修饰的片段,经过核酸外切酶酶切切掉互补序列,精确控制连接末端,利用基因重组原理不断延长链

长,得到越来越大片段的 DNA,最后得到全基因组 DNA。

本研究中的 DNA 组装技术,一方面,不需要 DNA 聚合酶修补核酸外切酶切割的不定数的缺口,核酸外切酶在硫修饰的位置停止切割,精确控制待连接末端的位置,不需要使用 DNA 聚合酶,减少 DNA 复制时发生突变的几率,降低了成本;另一方面,本研究利用 DNA 上的硫修饰与识别硫原子结构域(SBD)的靶向系统,招募 DNA 连接酶到待连接 DNA 末端进行大片段 DNA 的连接,可能进一步提高连接效率。可以将硫修饰依赖的 DNA 连接方法与 Gibson 组装法相结合,设计对应的硫修饰引物,将 Gibson 组装中的 *Taq* DNA 连接酶与 SBD 蛋白相融合,进一步提高 Gibson 方法的连接效率与可连接片段的长度。

随着合成生物学的发展,组装人造细胞的研究吸引着越来越多的科研工作者,DNA 组装技术是基因组合成的关键技术。面对如此长的基因组,组装技术的革新、效率的提升都能够让我们离探索生命的起源更进一步。

### REFERENCES

- [1] GIBSON DG, BENDERS GA, ANDREWS-PFANNKOCH C, DENISOVA EA, BADEN-TILLSON H, ZAVERI J, STOCKWELL TB, BROWNLEY A, THOMAS DW, ALGIRE MA, MERRYMAN C, YOUNG L, NOSKOV VN, GLASS JI, VENTER JC, HUTCHISON 3rd CA, SMITH HO. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome[J]. *Science*, 2008, 319(5867): 1215-1220.
- [2] GIBSON DG, BENDERS GA, AXELROD KC, ZAVERI J, ALGIRE MA, MOODIE M, MONTAGUE MG, VENTER JC, SMITH HO, HUTCHISON CA. One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome[J]. *Proceedings of National Academy of Sciences of the United States of America*, 2008,

- 105(51): 20404-20409.
- [3] GIBSON DG, GLASS JI, LARTIGUE C, NOSKOV VN, CHUANG RY, ALGIRE MA, BENDERS GA, MONTAGUE MG, MA L, MOODIE MM, MERRYMAN C, VASHEE S, KRISHNAKUMAR R, ASSAD-GARCIA N, ANDREWS-PFANNKOCHE C, DENISOVA EA, YOUNG L, QI ZQ, SEGALL-SHAPIRO TH, CALVEY CH, et al. Creation of a bacterial cell controlled by a chemically synthesized genome[J]. *Science*, 2010, 329(5987): 52-56.
- [4] ZHOU XF, HE XY, LIANG JD, LI AY, XU TG, KIESER T, HELMANN JD, DENG ZX. A novel DNA modification by sulphur[J]. *Molecular Microbiology*, 2005, 57(5): 1428-1438.
- [5] WANG LR, CHEN S, XU TG, TAGHIZADEH K, WISHNOK JS, ZHOU XF, YOU DL, DENG ZX, DEDON PC. Phosphorothioation of DNA in bacteria by dnd genes[J]. *Nature Chemical Biology*, 2007, 3(11): 709-710.
- [6] ARITA K, ARIYOSHI M, TOCHIO H, NAKAMURA Y, SHIRAKAWA M. Recognition of hemi-methylated DNA by the SRA protein UHRF<sub>1</sub> by a base-flipping mechanism[J]. *Nature*, 2008, 455(7214): 818-821.
- [7] AVVAKUMOV GV, WALKER JR, XUE S, LI YJ, DUAN SL, BRONNER C, ARROWSMITH CH, DHE-PAGANON S. Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF<sub>1</sub>[J]. *Nature*, 2008, 455(7214): 822-825.
- [8] HASHIMOTO H, HORTON JR, ZHANG X, BOSTICK M, JACOBSEN SE, CHENG XD. The SRA domain of UHRF<sub>1</sub> flips 5-methylcytosine out of the DNA helix[J]. *Nature*, 2008, 455(7214): 826-829.
- [9] RAJAKUMARA E, LAW JA, SIMANSHU DK, VOIGT P, JOHNSON LM, REINBERG D, PATEL DJ, JACOBSEN SE. A dual flip-out mechanism for 5mC recognition by the *Arabidopsis* SUVH<sub>5</sub> SRA domain and its impact on DNA methylation and H3K9 dimethylation *in vivo*[J]. *Genes & Development*, 2011, 25(2): 137-152.
- [10] HORTON JR, MABUCHI MY, COHEN-KARNI D, ZHANG X, GRIGGS RM, SAMARANAYAKE M, ROBERTS RJ, ZHENG Y, CHENG XD. Structure and cleavage activity of the tetrameric MspJI DNA modification-dependent restriction endonuclease[J]. *Nucleic Acids Research*, 2012, 40(19): 9763-9773.
- [11] HORTON JR, WANG H, MABUCHI MY, ZHANG X, ROBERTS RJ, ZHENG Y, WILSON GG, CHENG XD. Modification-dependent restriction endonuclease, MspJI, flips 5-methylcytosine out of the DNA helix[J]. *Nucleic Acids Research*, 2014, 42(19): 12092-12101.
- [12] HORTON JR, BORGARO JG, GRIGGS RM, QUIMBY A, GUAN SX, ZHANG X, WILSON GG, ZHENG Y, ZHU ZY, CHENG XD. Structure of 5-hydroxymethylcytosine-specific restriction enzyme, AbaSI, in complex with DNA[J]. *Nucleic Acids Research*, 2014, 42(12): 7947-7959.
- [13] KAZRANI AA, KOWALSKA M, CZAPINSKA H, BOCHTLER M. Crystal structure of the 5hmC specific endonuclease PvuRts1I[J]. *Nucleic Acids Research*, 2014, 42(9): 5929-5936.
- [14] LIU G, FU WC, ZHANG ZY, HE Y, YU H, WANG YL, WANG XL, ZHAO YL, DENG ZX, WU G, HE XY. Structural basis for the recognition of sulfur in phosphorothioated DNA[J]. *Nature Communications*, 2018, 9: 4689.
- [15] YU H, LI JY, LIU G, ZHAO G, WANG YL, HU WY, DENG ZX, WU G, GAN JH, ZHAO YL, HE XY. DNA backbone interactions impact the sequence specificity of DNA sulfur-binding domains: revelations from structural analyses[J]. *Nucleic Acids Research*, 2020, 48(15): 8755-8766.